

Лекция по эконометрике № 9

3 модуль

Модели множественного выбора (дополнительная тема)

Демидова

Ольга Анатольевна

https://www.hse.ru/staff/demidova_olga

E-mail: demidova@hse.ru

15.03.2021

- **Модели упорядоченного множественного выбора**
- **Мультиномиальные модели**

Напоминание о моделях бинарного выбора

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i,$$

$E(\varepsilon_i) = 0$, $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2$, $i = 1, \dots, n$,
 F – cumulative function of $\varepsilon/\sigma_\varepsilon$,
 $f = F'$ is a symmetric function.

$$\begin{cases} Y_i = 1, & \text{if } Y_i^* \geq 0 \\ Y_i = 0, & \text{if } Y_i^* < 0 \end{cases}$$

Отношение шансов (odd ratio) для логит модели

$$OR = \frac{\Pr(Y = 1)}{\Pr(Y = 0)}$$

$$\ln(OR) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Если X_j изменится на 1 то OR изменится в $\exp(\beta_j)$ раз.

Пример 1. Ответ на вопрос: «Насколько Вы доверяете правительству страны?»

Варианты ответов:

- 1 – полностью доверяю**
- 2 – скорее доверяю**
- 3 – скорее не доверяю**
- 4 – полностью не доверяю**

Пример 2. Состояние здоровья

Варианты ответов:

- 1 – fair
- 2 – good
- 3- excellent

Пример 3. Состояние на рынке труда

Варианты ответов:

- 1 – не работает
- 2 – частичная занятость (неполный рабочий день)
- 3 – полная занятость (полный рабочий день)

$$Y_i = 1, 2, \dots, m$$

$$Y_i^* = x_i' \beta + \varepsilon_i,$$

$$Y_i = j \quad \text{if} \quad c_{j-1} < Y_i^* < c_j, \quad j = 1, \dots, m,$$

$$c_0 = -\infty, \quad \dots, \quad c_m = \infty,$$

$$P(Y_i = j) = F(c_j - x_i' \beta) - F(c_{j-1} - x_i' \beta),$$

$$L = \prod_{j=1}^m \prod_{i: Y_i=j} (F(c_j - x_i' \beta) - F(c_{j-1} - x_i' \beta)) \rightarrow \max_{\beta, c}$$

Оценка параметров с помощью метода максимального правдоподобия

Проверка гипотезы о параллельности

$$P(Y_i = j) = F(c_j - x_i' \beta) - F(c_{j-1} - x_i' \beta),$$

$$\Rightarrow P(Y_i \leq k | X) = F(c_k - x_i' \beta), \quad k = 1, \dots, m$$

(parallel regression assumption)

Brant test

**Детали теста Бранта можно найти в W.Greene, 7
edition, p. 791**

Отношение шансов для упорядоченной логит модели

$$\frac{P(Y_i \leq k|X)}{P(Y_i > k|X)} = \exp(c_k - x_i' \beta), \quad k = 1, \dots, m$$

$$\Rightarrow \frac{P(Y_i \leq k|X)(X, x_j + 1)}{P(Y_i > k|X)(X, x_j + 1)} = \exp(-\beta_j)$$

$$\frac{P(Y_i > k|X)(X, x_j + 1)}{P(Y_i \leq k|X)(X, x_j + 1)} = \exp(\beta_j)$$

Предельные эффекты

$$\begin{aligned}\frac{\partial P(Y_i = 1)}{\partial X_k} &= -\beta_k f(c_1 - (X\beta)), \\ \frac{\partial P(Y = j)}{\partial X_k} &= -\beta_k [f(c_j - (X\beta)) - \\ &\quad - f(c_{j-1} - (X\beta))], \quad j = 1, \dots, m-1 \\ \Rightarrow \frac{\partial P(Y = j)}{\partial X_k} &= \beta_k f(c_{m-1} - (X\beta))\end{aligned}$$

Пример оценки моделей упорядоченного множественного выбора

Пример позаимствован у Ani Katchova,
<https://www.youtube.com/watch?v=c9kvqeLFF8U>

Зависимая переменная: $Y = \text{healthstatus}$

Значения: 1 – fair, 2 – good, 3- excellent

tab healthstatus

health status (fair, good, excellent)

	Freq.	Percent	Cum.
Fair	523	9.38	9.38
good	2,034	36.49	45.87
Excellent	3,017	54.13	100.00
Total	5,574	100.00	

Независимые переменные

sum age logincome numberdiseases

Variable	Obs	Mean	Std. Dev.	Min	Max
age	5,574	25.57613	16.73011	.0253251	63.27515
logincome	5,574	8.696929	1.220592	0	10.28324
numberdiseases	5,574	11.20526	6.788959	0	58.6

Результаты оценки

```
. ologit healthstatus age logincome numberdiseases
```

```
Iteration 0:  log likelihood = -5140.0463
Iteration 1:  log likelihood = -4776.008
Iteration 2:  log likelihood = -4769.8693
Iteration 3:  log likelihood = -4769.8525
Iteration 4:  log likelihood = -4769.8525
```

Ordered logistic regression

```
Number of obs      =      5,574
LR chi2(3)         =      740.39
Prob > chi2        =      0.0000
Pseudo R2         =      0.0720
```

Log likelihood = -4769.8525

healthstatus	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.0292944	.001681	-17.43	0.000	-.0325891	-.0259996
logincome	.2836537	.0231098	12.27	0.000	.2383593	.3289481
numberdiseases	-.0549905	.0040692	-13.51	0.000	-.0629661	-.047015
/cut1	-1.39598	.2061301			-1.799987	-.9919722
/cut2	.9513097	.2054301			.5486741	1.353945

Предельные эффекты

```
. margins, dydx(*) atmeans predict(outcome(1))
```

Conditional marginal effects

Number of obs = 5,574

Model VCE : OIM

Expression : Pr(healthstatus==1), predict(outcome(1))

dy/dx w.r.t. : age logincome numberdiseases

at : age = 25.57613 (mean)
logincome = 8.696929 (mean)
numberdiseases = 11.20526 (mean)

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.002058	.0001333	15.44	0.000	.0017969	.0023192
logincome	-.0199278	.0017344	-11.49	0.000	-.0233272	-.0165284
numberdiseases	.0038633	.0003056	12.64	0.000	.0032643	.0044623

Предельные эффекты

```
. margins, dydx(*) atmeans predict(outcome(2))
```

```
Conditional marginal effects      Number of obs      =      5,574
Model VCE      : OIM
```

```
Expression      : Pr(healthstatus==2), predict(outcome(2))
dy/dx w.r.t.    : age logincome numberdiseases
at              : age          =    25.57613 (mean)
                  logincome    =    8.696929 (mean)
                  numberdise~s =   11.20526 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0052244	.0003258	16.04	0.000	.0045859	.0058629
logincome	-.0505872	.0043054	-11.75	0.000	-.0590256	-.0421489
numberdiseases	.0098071	.000768	12.77	0.000	.0083018	.0113124

Предельные эффекты

```
. margins, dydx(*) atmeans predict(outcome(3))
```

Conditional marginal effects

Number of obs = 5,574

Model VCE : OIM

Expression : Pr(healthstatus==3), predict(outcome(3))

dy/dx w.r.t. : age logincome numberdiseases

at : age = 25.57613 (mean)
logincome = 8.696929 (mean)
numberdise~s = 11.20526 (mean)

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.0072824	.0004179	-17.43	0.000	-.0081014	-.0064634
logincome	.070515	.0057527	12.26	0.000	.05924	.0817901
numberdiseases	-.0136704	.0010126	-13.50	0.000	-.015655	-.0116858

Предсказанные значения зависимой переменной

```
. predict p1ologit p2ologit p3ologit, pr
```

```
. summarize p1ologit p2ologit p3ologit
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p1ologit	5,574	.0946903	.0843148	.0233629	.859022
p2ologit	5,574	.3651672	.0946158	.1255265	.5276064
p3ologit	5,574	.5401425	.1640575	.0154515	.7999009

```
. tab healthstatus
```

health status (fair, good, excellent)	Freq.	Percent	Cum.
1	523	9.38	9.38
2	2,034	36.49	45.87
3	3,017	54.13	100.00
Total	5,574	100.00	

Пример 1. Отношение шансов

```
. ologit healthstatus age logincome numberdiseases, or
```

```
Iteration 0:  log likelihood = -5140.0463
Iteration 1:  log likelihood = -4776.008
Iteration 2:  log likelihood = -4769.8693
Iteration 3:  log likelihood = -4769.8525
Iteration 4:  log likelihood = -4769.8525
```

Ordered logistic regression

```
Number of obs      =      5,574
LR chi2(3)         =      740.39
Prob > chi2        =      0.0000
Pseudo R2         =      0.0720
```

Log likelihood = -4769.8525

healthstatus	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.9711306	.0016325	-17.43	0.000	.9679362	.9743355
logincome	1.327973	.0306892	12.27	0.000	1.269165	1.389506
numberdiseases	.9464941	.0038515	-13.51	0.000	.9389753	.9540731
/cut1	-1.39598	.2061301			-1.799987	-.9919722
/cut2	.9513097	.2054301			.5486741	1.353945

Пример 1. Тест Бранта

```
. brant
```

Brant Test of Parallel Regression Assumption

Variable	chi2	p>chi2	df
All	17.45	0.001	3
age	0.18	0.672	1
logincome	17.40	0.000	1
numberdis~s	0.20	0.659	1

A significant test statistic provides evidence that the parallel regression assumption has been violated.

Пример 1. Сфера деятельности

Варианты ответов:

- 1 – бизнес
- 2 – образование
- 3 – некоммерческая организация

Пример 2. Способ рыбалки

Варианты ответов:

- 1 – с берега
- 2 – с причала
- 3 – с собственной лодки
- 4 – с арендованной лодки

Зависимая переменная не является упорядоченной,
Существует выбор между альтернативами $j = 1, 2, \dots, M$,
 U_{ij} — полезность j -ой альтернативы для i -го индивида.

$$P\{Y_i = j\} = P\{U_{ij} = \max\{U_{i1}, \dots, U_{iM}\}\},$$
$$U_{ij} = x'_i \beta_j + \varepsilon_{ij}.$$

Задача допускает аналитическое решение, если ε_{ij}
независимы и имеют функцию распределения
 $F(x) = \exp\{-\exp(-x)\}$

$$P(Y_i = 1) = \frac{1}{1 + \exp(x'_i \beta_2) + \dots + \exp(x'_i \beta_m)},$$

$$P(Y_i = j) = \frac{\exp(x'_i \beta_j)}{1 + \exp(x'_i \beta_2) + \dots + \exp(x'_i \beta_m)}$$

$$\Rightarrow \frac{P(Y_i = j)}{P(Y_i = k)} = \frac{\exp(x'_i \beta_j)}{\exp(x'_i \beta_k)} = \exp(x'_i (\beta_j - \beta_k))$$

IIA – independence from irrelevant alternatives
Test Small-Hsiao

Пример 1

Пример 1 позаимствован у Ani Katchova,
<https://www.youtube.com/watch?v=iqypob4My4o>

Зависимая переменная $Y = \text{mode}$ = Способ рыбалки

Значения:

- 1 – beach (с берега)
- 2 – pier (с причала)
- 3 – private (с собственной лодки)
- 4 – charter (с арендованной лодки)

Пример

```
. mlogit mode income
```

```
Iteration 0: log likelihood = -1497.7229
Iteration 1: log likelihood = -1477.5265
Iteration 2: log likelihood = -1477.1514
Iteration 3: log likelihood = -1477.1506
Iteration 4: log likelihood = -1477.1506
```

Multinomial logistic regression

```
Number of obs      =      1,182
LR chi2(3)         =      41.14
Prob > chi2        =      0.0000
Pseudo R2         =      0.0137
```

Log likelihood = -1477.1506

mode	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
beach						
income	.0316399	.0418463	0.76	0.450	-.0503774	.1136571
_cons	-1.341291	.1945167	-6.90	0.000	-1.722537	-.9600457
pier						
income	-.111763	.0439795	-2.54	0.011	-.1979612	-.0255649
_cons	-.5271412	.1777842	-2.97	0.003	-.8755918	-.1786906
private						
income	.1235462	.0279106	4.43	0.000	.0688425	.17825
_cons	-.6023707	.1360964	-4.43	0.000	-.8691147	-.3356267
charter	(base outcome)					

Пример

```
. mlogit mode income, baseoutcome(2)
```

```
Iteration 0: log likelihood = -1497.7229
Iteration 1: log likelihood = -1477.5265
Iteration 2: log likelihood = -1477.1514
Iteration 3: log likelihood = -1477.1506
Iteration 4: log likelihood = -1477.1506
```

Multinomial logistic regression

```
Number of obs      =      1,182
LR chi2(3)         =      41.14
Prob > chi2        =      0.0000
Pseudo R2         =      0.0137
```

Log likelihood = -1477.1506

mode	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
beach						
income	.1434029	.0532884	2.69	0.007	.0389595	.2478463
_cons	-.8141503	.228632	-3.56	0.000	-1.262261	-.3660399
pier	(base outcome)					
private						
income	.2353093	.0436681	5.39	0.000	.1497214	.3208971
_cons	-.0752295	.1832396	-0.41	0.681	-.4343724	.2839134
charter						
income	.111763	.0439795	2.54	0.011	.0255649	.1979612
_cons	.5271412	.1777842	2.97	0.003	.1786906	.8755918



```
Conditional marginal effects      Number of obs      =      1,182
Model VCE      : OIM

Expression      : Pr(mode==beach), predict(pr outcome(1))
dy/dx w.r.t.   : income
at              : income          =      4.099337 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
income	.000075	.0039337	0.02	0.985	-.0076349	.0077848

```
. margins, dydx(*) atmeans predict(pr outcome(2))

Conditional marginal effects      Number of obs      =      1,182
Model VCE      : OIM

Expression      : Pr(mode==pier), predict(pr outcome(2))
dy/dx w.r.t.    : income
at              : income          =      4.099337 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
income	-.0206598	.0048735	-4.24	0.000	-.0302117	-.011108

Предельные эффекты

```
. margins, dydx(*) atmeans predict(pr outcome(3))
```

```
Conditional marginal effects      Number of obs      =      1,182
Model VCE      : OIM
```

```
Expression      : Pr(mode==private), predict(pr outcome(3))
dy/dx w.r.t.    : income
at              : income          =      4.099337 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
income	.0325985	.005692	5.73	0.000	.0214424	.0437547

```
.
. margins, dydx(*) atmeans predict(pr outcome(4))
```

```
Conditional marginal effects      Number of obs      =      1,182
Model VCE      : OIM
```

```
Expression      : Pr(mode==charter), predict(pr outcome(4))
dy/dx w.r.t.    : income
at              : income          =      4.099337 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
income	-.0120137	.0060756	-1.98	0.048	-.0239215	-.0001058

Предсказанные значения зависимой переменной

```
. predict pmlogit1 pmlogit2 pmlogit3 pmlogit4, pr
```

```
.
```

```
. summarize pmlogit1 pmlogit2 pmlogit3 pmlogit4
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pmlogit1	1,182	.1133672	.0036716	.0947395	.1153659
pmlogit2	1,182	.1505922	.0444575	.0356142	.2342903
pmlogit3	1,182	.3536379	.0797714	.2396973	.625706
pmlogit4	1,182	.3824027	.0346281	.2439403	.4158273

```
.
```

```
. tab mode
```

Fishing mode	Freq.	Percent	Cum.
beach	134	11.34	11.34
pier	178	15.06	26.40
private	418	35.36	61.76
charter	452	38.24	100.00
Total	1,182	100.00	

Пример 2. Продолжение примера про здоровье

```
. mlogit healthstatus age logincome numberdiseases
```

```
Iteration 0:    log likelihood = -5140.0463
Iteration 1:    log likelihood = -4826.7356
Iteration 2:    log likelihood = -4772.0141
Iteration 3:    log likelihood = -4766.2058
Iteration 4:    log likelihood = -4766.1902
Iteration 5:    log likelihood = -4766.1902
```

Multinomial logistic regression

```
Number of obs      =      5,574
LR chi2(6)         =      747.71
Prob > chi2        =      0.0000
Pseudo R2         =      0.0727
```

Log likelihood = -4766.1902

healthstatus	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fair						
age	.0424728	.0031346	13.55	0.000	.0363292	.0486165
logincome	-.4191076	.0323039	-12.97	0.000	-.4824221	-.3557931
numberdiseases	.0833499	.006825	12.21	0.000	.0699731	.0967266
_cons	-.3537368	.282676	-1.25	0.211	-.9077716	.2002979
good						
age	.025483	.0018533	13.75	0.000	.0218506	.0291155
logincome	-.1473865	.0291931	-5.05	0.000	-.2046039	-.0901692
numberdiseases	.0454721	.004827	9.42	0.000	.0360114	.0549328
_cons	-.2366722	.2592993	-0.91	0.361	-.7448895	.2715451
excellent	(base outcome)					

Пример 2. Продолжение примера про здоровье

```
. mlogit healthstatus age logincome numberdiseases, baseoutcome(2)
```

```
Iteration 0:   log likelihood = -5140.0463
Iteration 1:   log likelihood = -4826.7356
Iteration 2:   log likelihood = -4772.0141
Iteration 3:   log likelihood = -4766.2058
Iteration 4:   log likelihood = -4766.1902
Iteration 5:   log likelihood = -4766.1902
```

Multinomial logistic regression

```
Number of obs      =      5,574
LR chi2(6)         =      747.71
Prob > chi2        =      0.0000
Pseudo R2         =      0.0727
```

Log likelihood = -4766.1902

healthstatus	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fair						
age	.0169898	.0031035	5.47	0.000	.010907	.0230726
logincome	-.2717211	.0286096	-9.50	0.000	-.327795	-.2156473
numberdiseases	.0378778	.0062911	6.02	0.000	.0255475	.050208
_cons	-.1170646	.2524238	-0.46	0.643	-.6118062	.3776769
good	(base outcome)					
excellent						
age	-.025483	.0018533	-13.75	0.000	-.0291155	-.0218506
logincome	.1473865	.0291931	5.05	0.000	.0901692	.2046039
numberdiseases	-.0454721	.004827	-9.42	0.000	-.0549328	-.0360114
_cons	.2366722	.2592993	0.91	0.361	-.2715451	.7448895

Пример 2. Предельные эффекты для исхода 1

```
. margins, dydx(*) atmeans predict(pr outcome(1))
```

```
Conditional marginal effects      Number of obs      =      5,574
Model VCE      : OIM
```

```
Expression      : Pr(healthstatus==fair), predict(pr outcome(1))
dy/dx w.r.t.    : age logincome numberdiseases
at              : age              =      25.57613 (mean)
                  logincome        =      8.696929 (mean)
                  numberdise~s     =      11.20526 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0022253	.0001985	11.21	0.000	.0018363	.0026144
logincome	-.0249397	.0019742	-12.63	0.000	-.0288091	-.0210703
numberdiseases	.004497	.0004403	10.21	0.000	.003634	.0053601

Пример 2. Сравнение предельных эффектов ologit и mlogit для исхода 1

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.002058	.0001333	15.44	0.000	.0017969	.0023192
logincome	-.0199278	.0017344	-11.49	0.000	-.0233272	-.0165284
numberdiseases	.0038633	.0003056	12.64	0.000	.0032643	.0044623

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0022253	.0001985	11.21	0.000	.0018363	.0026144
logincome	-.0249397	.0019742	-12.63	0.000	-.0288091	-.0210703
numberdiseases	.004497	.0004403	10.21	0.000	.003634	.0053601

Пример 2. Сравнение предельных эффектов ologit и mlogit для исхода 2

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0052244	.0003258	16.04	0.000	.0045859	.0058629
logincome	-.0505872	.0043054	-11.75	0.000	-.0590256	-.0421489
numberdiseases	.0098071	.000768	12.77	0.000	.0083018	.0113124

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0047914	.0004161	11.52	0.000	.0039759	.0056069
logincome	-.022744	.0064079	-3.55	0.000	-.0353032	-.0101848
numberdiseases	.0083332	.0010648	7.83	0.000	.0062462	.0104203

Пример 2. Сравнение предельных эффектов ologit и mlogit для исхода 3

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.0072824	.0004179	-17.43	0.000	-.0081014	-.0064634
logincome	.070515	.0057527	12.26	0.000	.05924	.0817901
numberdiseases	-.0136704	.0010126	-13.50	0.000	-.015655	-.0116858

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.0070167	.000441	-15.91	0.000	-.0078811	-.0061524
logincome	.0476837	.006889	6.92	0.000	.0341815	.0611859
numberdiseases	-.0128302	.0011574	-11.09	0.000	-.0150987	-.0105618

Пример 2. Предсказанные значения зависимой переменной

```
. predict pmlogit1 pmlogit2 pmlogit3, pr
```

```
. summarize pmlogit1 pmlogit2 pmlogit3
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pmlogit1	5,574	.0938285	.0903648	.0168382	.8635994
pmlogit2	5,574	.3649085	.0860006	.1258262	.5486323
pmlogit3	5,574	.541263	.1576133	.0105744	.7683809

```
. tab healthstatus
```

health status (fair, good, excellent)	Freq.	Percent	Cum.
fair	523	9.38	9.38
good	2,034	36.49	45.87
excellent	3,017	54.13	100.00
Total	5,574	100.00	



NATIONAL RESEARCH
UNIVERSITY

Thank you for your attention!

20, Myasnitskaya str., Moscow, Russia, 101000
Tel.: +7 (495) 628-8829, Fax: +7 (495) 628-7931
www.hse.ru